WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

# CONDITIONAL PREDICTION AND UNBIASEDNESS IN STRUCTURAL EQUATIONS

Gordon M. Kaufman

12 August 1965

138-65

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

# CONDITIONAL PREDICTION AND UNBIASEDNESS IN STRUCTURAL EQUATIONS

Gordon M. Kaufman

12 August 1965

138-65

CONDITIONAL PREDICTION AND UNBIASEDNESS

IN STRUCTURAL EQUATIONS

Gordon M. Kaufman

12 August 1965

## 1. Introduction

Sewall [ 4 ], Waugh [ 6 ], and Srinivasan [ 5 ] discuss least squares estimators of a very particular set of two structural equations with no exogenous variables and the net result of their discussion is that such estimators are unbiased predictors of one dependent variable given the other. Our purpose here is to generalize Sewall's main result in several directions, and to provide some ancillary facts about the structural equations defined in (2.1) below. Throughout this note we distinguish a random variable from a value assumed by it with a tilde; e.g. the random matrix $\tilde{\underline{\underline{\epsilon}}}$. The symbol $\otimes$ denotes the Kroenecker direct product; e.g. $\underline{\underline{A}} \otimes \underline{\underline{B}}$. And we say that the $(m \times m)$ random matrix $\tilde{\underline{\underline{\epsilon}}}$ is "Wishart with parameter set $(\underline{\underline{H}}, n)$" if $\tilde{\underline{\underline{\epsilon}}}$ has density

$$\begin{cases} c_o e^{-\frac{1}{2}\mathrm{tr}\, \underline{\underline{H}}\, \underline{\underline{\epsilon}}} \, |\underline{\underline{\epsilon}}|^{\frac{1}{2}(n-m-1)} & , \quad \underline{\underline{\epsilon}}, \underline{\underline{H}} \text{ PDS}, \; n > m-1 , \\ 0 & \text{otherwise} \end{cases}$$

where $c_o$ is a normalizing constant.

## 2. Generalization

Two questions immediately arise: First, "What is the analogue of Sewell's assertion about $\underline{x}^t\underline{y}/\underline{x}^t\underline{x}$ when $m > 2$, $\underline{\underline{B}}$ is arbitrary $(m \times m)$ non-singular, and exogenous variables are present?" And second, "What does the answer to the first question imply about the conditional expectation of any single endoge-

nous variable given the values of all other endogenous variables and the values of the exogenous variables?"

To make these questions precise, consider the following system of stochastic equations:

$$\underline{\underline{B}}\, \tilde{\underline{y}}^{(j)} = -\underline{\underline{\Gamma}}\, \underline{z}^{(j)} + \tilde{\underline{u}}^{(j)} \tag{2.1}$$

where $\underline{\underline{B}}$ and $\underline{\underline{\Gamma}}$ are (m x m) and (m x r) coefficient matrices, fixed for all j, $\underline{z}^{(j)}$ is an (r x 1) vector of predetermined variables and $\tilde{\underline{y}}^{(j)}$ and $\tilde{\underline{u}}^{(j)}$ are (m x 1) random vectors. We assume that $\{\tilde{\underline{u}}^{(j)}$, $j=1,2,\ldots\}$ is a sequence of mutually independent, identically Normal random vectors with mean $\underline{0}$ and PDS covariance matrix $\underline{\underline{\Sigma}} \equiv \underline{\underline{h}}^{-1}$; and $\underline{\underline{B}}$ is non-singular. One observes $(\underline{y}^{(j)}, \underline{z}^{(j)})$ $j=1,2,\ldots$ but neither $\underline{\underline{B}}$, nor $\underline{\underline{\Gamma}}$ nor $\underline{\underline{h}}$ is known with certainty. If we partition a generic $\underline{y}$ into $\underline{y} = (\underline{y}_1^t\ \underline{y}_2^t)^t$ with $\underline{y}_1$ of dimension (p x 1), $1 \leq p \leq m-1$, conformably partition $\underline{\underline{\Pi}} \equiv -\underline{\underline{B}}^{-1}\,\underline{\underline{\Gamma}}$ into $\underline{\underline{\Pi}} = \begin{bmatrix} \underline{\underline{\Pi}}_1 \\ \underline{\underline{\Pi}}_2 \end{bmatrix}$, $\underline{\underline{\Pi}}_1$ of dimension (p x r), let $\underline{\underline{\Omega}} = \underline{\underline{B}}^{-1}\,\underline{\underline{\Sigma}}\,\underline{\underline{B}}^{-1t}$ and set

$$\underline{\underline{\Omega}}^{-1} = \begin{bmatrix} \underline{\underline{\Omega}}_{11} & \underline{\underline{\Omega}}_{12} \\ \underline{\underline{\Omega}}_{21} & \underline{\underline{\Omega}}_{22} \end{bmatrix}^{-1} = \underline{\underline{H}} \begin{bmatrix} \underline{\underline{H}}_{11} & \underline{\underline{H}}_{12} \\ \underline{\underline{H}}_{21} & \underline{\underline{H}}_{22} \end{bmatrix}, \quad \underline{\underline{\Omega}}_{11},\ \underline{\underline{H}}_{11} \text{ dim (p x p)}$$

then

$$E(\tilde{\underline{y}}_1 | \underline{y}_2, \underline{z}) = \underline{\underline{\Pi}}_1\, \underline{z} - \underline{\underline{H}}_{11}^{-1}\, \underline{\underline{H}}_{21} [\underline{y}_2 - \underline{\underline{\Pi}}_2\, \underline{z}] \tag{2.2}$$

and

$$\mathrm{Var}(\tilde{\underline{y}}_1 | \tilde{\underline{y}}_2, \underline{z}) = \underline{\underline{H}}_{11}^{-1} = [\underline{\underline{\Omega}}_{11} - \underline{\underline{\Omega}}_{12}\, \underline{\underline{\Omega}}_{22}^{-1}\, \underline{\underline{\Omega}}_{21}] \quad . \tag{2.3}$$

Now suppose we observe a sequence $\{(\underline{y}^{(j)}, \underline{z}^{(j)}), j=1,2,\ldots,n\}$ of $n \geq r+m$ sample observations generated according to (2.2). Let

$$\underline{\underline{Y}} = [\underline{y}^{(1)},\ldots,\underline{y}^{(n)}] \quad , \quad \underline{\underline{Z}} = [\underline{z}^{(1)},\ldots,\underline{z}^{(n)}] \quad , \quad \underline{\underline{Z}} \text{ of rank } r,$$

and

$$\underline{\underline{V}} = \Sigma\, \underline{z}^{(j)}\, \underline{z}^{(j)^t} = \underline{\underline{Z}}\, \underline{\underline{Z}}^t \quad .$$

Then it is well known ([1] p. 183) that given $\underline{\underline{B}}$, $-\underline{\underline{\Gamma}}$, and $\underline{\underline{\Sigma}}$, the statistics

$$\tilde{\underline{\underline{P}}} \equiv \tilde{\underline{\underline{Y}}}\, \underline{\underline{Z}}^t\, \underline{\underline{V}}^{-1} \quad \text{and} \quad \tilde{\underline{\underline{\varepsilon}}} \equiv \Sigma[\tilde{\underline{y}}^{(j)} - \tilde{\underline{\underline{P}}}\, \underline{z}^{(j)}]\, [\tilde{\underline{y}}^{(j)} - \tilde{\underline{\underline{P}}}\, \underline{z}^{(j)}]^t$$

are mutually independent and that $\underline{\underline{P}}$ is an unbiased estimator of $\underline{\underline{\Pi}}$ and $\frac{1}{n-r}\,\underline{\underline{\varepsilon}}$ an unbiased estimator of $\underline{\underline{\Omega}}$. Partition $\underline{\underline{P}}$ and $\underline{\underline{\varepsilon}}$ as follows:

$$\underline{\underline{P}} = \begin{bmatrix} \underline{\underline{P}}_1 \\ \underline{\underline{P}}_2 \end{bmatrix} \quad , \quad \underline{\underline{P}}_1 \text{ dim } (p \times r)$$

and

$$\underline{\underline{\varepsilon}} = \begin{bmatrix} \underline{\underline{\varepsilon}}_{11} & \underline{\underline{\varepsilon}}_{12} \\ \underline{\underline{\varepsilon}}_{21} & \underline{\underline{\varepsilon}}_{22} \end{bmatrix} \quad , \quad \underline{\underline{\varepsilon}}_{11} \text{ dim } (p \times p) \quad .$$

The analogue of the central question of [ 4 ], [ 5 ], and [ 6 ] may now be stated: Is it true that

$$E(\tilde{\underline{y}}_1^{(n+1)} | \underline{y}_2^{(n+1)}, \underline{z}^{(n+1)}) = E(\tilde{\underline{\underline{P}}}_1)\, \underline{z}^{(n+1)} - E(\tilde{\underline{\underline{\varepsilon}}}_{12}\, \tilde{\underline{\underline{\varepsilon}}}_{22}^{-1})\, [\underline{y}_2^{(n+1)} - E(\tilde{\underline{\underline{P}}}_2)]\underline{z}^{(n+1)} ? \quad (2.4)$$

The answer is "yes" and this is proven shortly. We go on, however, and do considerably more in the next section.

Defining $\underline{\underline{\varepsilon}}_{11.2} = \underline{\underline{\varepsilon}}_{11} - \underline{\underline{\varepsilon}}_{12}\, \underline{\underline{\varepsilon}}_{22}^{-1}\, \underline{\underline{\varepsilon}}_{21}$ and $\underline{\underline{R}}_{1.2} = \underline{\underline{\varepsilon}}_{12}\, \underline{\underline{\varepsilon}}_{22}^{-1}$, we show that the joint likelihood of $(\tilde{\underline{\underline{P}}}, \tilde{\underline{\underline{R}}}_{1.2}, \tilde{\underline{\underline{\varepsilon}}}_{11.2}, \tilde{\underline{\underline{\varepsilon}}}_{22})$ given $\underline{\underline{\Pi}}$ and $\underline{\underline{\Omega}}$ has these properties:

(1) $\tilde{\underline{\underline{P}}}$ is Normal with mean $\underline{\underline{\Pi}}$ and variance-covariance matrix $\underline{\underline{H}} \otimes \underline{\underline{V}}$, and is independent of $\tilde{\underline{\underline{R}}}_{1.2}$, $\tilde{\underline{\underline{\varepsilon}}}_{11.2}$, and $\tilde{\underline{\underline{\varepsilon}}}_{22}$.

(2) $\tilde{\underline{\underline{\varepsilon}}}_{11.2}$ is (p x p) Wishart with parameter set $(\underline{\underline{H}}_{11}, n-r)$ and is independent of $\tilde{\underline{\underline{P}}}$, $\tilde{\underline{\underline{R}}}_{1.2}$, and $\tilde{\underline{\underline{\varepsilon}}}_{22}$; consequently $E(\tilde{\underline{\underline{\varepsilon}}}_{11.2}) = (n-r)\underline{\underline{H}}_{11}^{-1}$.

(3) $\tilde{\underline{\underline{R}}}_{1.2}$ and $\tilde{\underline{\underline{\varepsilon}}}_{22}$ are jointly independent of $\tilde{\underline{\underline{P}}}$ and $\tilde{\underline{\underline{\varepsilon}}}_{11.2}$, while the conditional density of $\tilde{\underline{\underline{R}}}_{1.2}$ given $\tilde{\underline{\underline{\varepsilon}}}_{22}$ is Normal with mean $\underline{\underline{H}}_{11}^{-1}\, \underline{\underline{H}}_{12}$ and variance-covariance matrix $\underline{\underline{H}}_{11} \otimes \underline{\underline{\varepsilon}}_{22}$ and the marginal density of $\tilde{\underline{\underline{\varepsilon}}}_{22}$ is Wishart with parameter set $(\underline{\underline{\Omega}}_{22}^{-1}, n+p-r)$. And $E(\tilde{\underline{\underline{\varepsilon}}}_{22}) = (n+p-r)\underline{\underline{\Omega}}_{22}$.

(4) The density of $\tilde{\underline{\underline{R}}}_{1.2}$ unconditional as regards $\tilde{\underline{\underline{P}}}$, $\tilde{\underline{\underline{\varepsilon}}}_{22}$, and $\tilde{\underline{\underline{\varepsilon}}}_{11.2}$ is

$$c''\left|[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\, \underline{\underline{H}}_{12}]\underline{\underline{\Omega}}_{22}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\, \underline{\underline{H}}_{12}]^{t} + \underline{\underline{H}}_{11}^{-1}\right|^{-\frac{1}{2}(n+p-r)} \tag{2.5}$$

where c" is a normalizing constant equal to

$$\frac{\Gamma_p(\frac{1}{2}[n+p-r])}{2^{\frac{1}{2}p(r-p)}\ \pi^{\frac{1}{2}pr}\ \Gamma_p(\frac{1}{2}[n-r])}\ |\underline{\underline{\Omega}}_{22}|^{\frac{1}{2}r}\ |\underline{\underline{H}}_{11}|^{-\frac{1}{2}(n-r)}$$

and $\Gamma_p(\alpha) = \prod_{i=1}^{p} \Gamma(\alpha-\frac{1}{2}(i-1))$. This is the generalized multivariate Student density first derived by Savage [ 3 ]. Here $\tilde{\underline{\underline{R}}}_{1.2}$ has mean $\underline{\underline{H}}_{11}^{-1}\, \underline{\underline{H}}_{12}$ and variance-covariance matrix $\frac{1}{n-r-1}\, \underline{\underline{\Omega}}_{22}^{-1} \otimes \underline{\underline{H}}_{11}^{-1}$. (See Martin [ 2 ]).

The above properties clearly imply that (2.4) holds and in addition, since $\mathrm{Var}(\tilde{y}_1^{(n+1)} | y_2^{(n+1)}, z^{(n+1)}) = H_{11}^{-1}$, imply that $\frac{1}{(n-r)} \varepsilon_{11.2}$ is an unbiased estimate of $\mathrm{Var}(\tilde{y}_1^{(n+1)} | y_2^{(n+1)}, z^{(n+1)})$. Similarly, since $\mathrm{Var}(\tilde{y}_2 | z) = \varepsilon_{22}$, an unbiased estimate of this variance-covariance matrix is $\frac{1}{n+p-r} \varepsilon_{22}$. Setting p=1 and running down the list of assertions above answers the second question posed at the outset of this section.

## 3. Proofs

It is well known (see [ 1 ], p. 183) that the joint likelihood of $(\underline{P}, \tilde{\varepsilon})$ given $\underline{V}$, and $\underline{H}$, is

$$c\, e^{-\frac{1}{2}\mathrm{tr}\, \underline{H}\{[\underline{P}-\underline{\Pi}]\underline{V}[\underline{P}-\underline{\Pi}]^t\}} \cdot e^{-\frac{1}{2}\mathrm{tr}\, \underline{H}\, \underline{\varepsilon}} \, |\underline{\varepsilon}|^{\frac{1}{2}(n-r-m-1)} \tag{3.1}$$

where c is a normalizing constant. To find the joint density of $(\tilde{P}, \tilde{R}_{1.2}, \tilde{\varepsilon}_{11.2}, \tilde{\varepsilon}_{22})$ we need the following

Lemma: The Jacobian $J(\underline{P}, \underline{\varepsilon} \to \underline{P}, \underline{R}_{1.2}, \underline{\varepsilon}_{11.2}, \underline{\varepsilon}_{22})$ of the transformation from $(\underline{P}, \underline{\varepsilon})$ to $(\underline{P}, \underline{R}_{1.2}, \underline{\varepsilon}_{11.2}, \underline{\varepsilon}_{22})$ is $|\underline{\varepsilon}_{22}|^p$.

Proof: We split the Jacobian into the product of four transformations done successively: $\underline{P} \to \underline{P}$, $\varepsilon_{22} \to \varepsilon_{22}$, $\varepsilon_{12} \to R_{1.2}$ and $\varepsilon_{11} \to \varepsilon_{11.2}$. First, $J(\underline{P} \to \underline{P}) = J(\varepsilon_{22} \to \varepsilon_{22}) = 1$. Now $R_{1.2} = \varepsilon_{12}\, \varepsilon_{22}^{-1}$, and a little algebra shows that $\varepsilon_{11} = \varepsilon_{11.2} + R_{1.2}\, \varepsilon_{22}\, R_{1.2}^t$. Consequently $J(\varepsilon_{12} \to R_{1.2}) = |\varepsilon_{22}|^p$ and $J(\varepsilon_{11} \to \varepsilon_{11.2}) = 1$. Multiplying all four Jacobians together proves the lemma.

We can now find the joint density of $(\tilde{P}, \tilde{R}_{1.2}, \tilde{\varepsilon}_{11.2}, \tilde{\varepsilon}_{22})$ from that of $(\tilde{P}, \tilde{\varepsilon})$ by substituting in (3.1) and multiplying by the Jacobian $|\varepsilon_{22}|^p$. First rewrite

The above properties clearly imply that (2.4) holds and in addition, since $\mathrm{Var}(\underline{Y}_1^{-(n+1)}|\underline{Y}_2^{(n+1)}, \underline{\Sigma}^{(n+1)}) = \underline{\Sigma}_{11.2}$, imply that $\frac{1}{(n-r)}\underline{S}_{11.2}$ is an unbiased estimate of $\mathrm{Var}(\underline{Y}_1^{-(n+1)}|\underline{Y}_2^{(n+1)}, \underline{\Sigma}^{(n+1)})$. Alternatively, since [illegible] an unbiased estimate of this variance-covariance matrix is $\frac{1}{n-p-1}\underline{S}_{11.2}$. Setting p=1 and running down the list of cases [illegible] answers the second question posed at the outset of this section.

3. Proofs

It is well known (see, e.g., [illegible]) that the joint likelihood of $\underline{Y}$, $\underline{\Sigma}$ given $\underline{Y}_1$ and $\underline{Y}_2$ [illegible]

$$[illegible] \quad (3.1)$$

where [illegible]

[illegible]

Lemma [illegible]

[illegible]

$$\mathrm{tr}\ \underline{\underline{H}}\ \underline{\underline{\varepsilon}} = \mathrm{tr}\ \{\underline{\underline{H}}_{11}\ \underline{\underline{\varepsilon}}_{11} + \underline{\underline{H}}_{12}\ \underline{\underline{\varepsilon}}_{21} + \underline{\underline{H}}_{21}\ \underline{\underline{\varepsilon}}_{12} + \underline{\underline{H}}_{22}\ \underline{\underline{\varepsilon}}_{22}\}$$

$$= \mathrm{tr}\ \underline{\underline{H}}_{11}\ \underline{\underline{\varepsilon}}_{11.2} + \mathrm{tr}\ \underline{\underline{\Omega}}_{22}^{-1}\ \underline{\underline{\varepsilon}}_{22} + \mathrm{tr}\ \underline{\underline{H}}_{11}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\underline{\underline{H}}_{12}]\underline{\underline{\varepsilon}}_{22}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\underline{\underline{H}}_{12}]^t$$

by writing out $\mathrm{tr}\ \underline{\underline{H}}\ \underline{\underline{\varepsilon}}$, substituting using the definitions of $\underline{\underline{\Omega}}_{22}$ in terms of the $\underline{\underline{H}}_{ij}$, $1 \leq i, j \leq 2$, and of $\underline{\underline{R}}_{1.2}$ and $\underline{\underline{\varepsilon}}_{11.2}$ in terms of the $\underline{\underline{\varepsilon}}_{ij}$, $i \leq i, j \leq 2$. The joint density (3.1) multiplied by $|\underline{\underline{\varepsilon}}|^p$ then gives that of $(\tilde{\underline{\underline{P}}}, \tilde{\underline{\underline{R}}}_{1.2}, \tilde{\underline{\underline{\varepsilon}}}_{11.2}, \tilde{\underline{\underline{\varepsilon}}}_{22})$ as

$$c'\ e^{-\frac{1}{2}\mathrm{tr}\ \underline{\underline{H}}\{[\underline{\underline{P}}-\underline{\underline{\Pi}}]\underline{\underline{V}}[\underline{\underline{P}}-\underline{\underline{\Pi}}]^t\}} \cdot e^{-\frac{1}{2}\mathrm{tr}\ \underline{\underline{H}}_{11}\ \underline{\underline{\varepsilon}}_{11.2}}\ |\underline{\underline{\varepsilon}}_{11.2}|^{\frac{1}{2}(n-r-m-1)}$$

$$\cdot\ e^{-\frac{1}{2}\mathrm{tr}\ \underline{\underline{\Omega}}_{22}^{-1}\ \underline{\underline{\varepsilon}}_{22}}\ |\underline{\underline{\varepsilon}}_{22}|^{\frac{1}{2}(n+p-r-m-1)} \qquad (3.2)$$

$$\cdot\ e^{-\frac{1}{2}\mathrm{tr}\ \underline{\underline{H}}_{11}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\ \underline{\underline{H}}_{12}]\underline{\underline{\varepsilon}}_{22}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\ \underline{\underline{H}}_{12}]^t}\ |\underline{\underline{\varepsilon}}_{22}|^{\frac{1}{2}p}$$

and the assertions 1, 2, and 3 of the previous section follow directly.

To prove assertion 4, rewrite that part of (3.2) involving $\underline{\underline{\varepsilon}}_{22}$ as

$$e^{-\frac{1}{2}\mathrm{tr}\ \underline{\underline{A}}\ \underline{\underline{\varepsilon}}_{22}}\ |\underline{\underline{\varepsilon}}_{22}|^{\frac{1}{2}(n+2p-r-m-1)}, \qquad (3.3)$$

where $\underline{\underline{A}} = [\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\ \underline{\underline{H}}_{12}]^t\ \underline{\underline{H}}_{11}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\ \underline{\underline{H}}_{12}] + \underline{\underline{\Omega}}_{22}^{-1}$, integrate with respect to $\underline{\underline{P}}$, $\underline{\underline{\varepsilon}}_{11.2}$, and $\underline{\underline{\varepsilon}}_{22}$, and use the fact that the constant that normalizes (3.3) is $c^*|\underline{\underline{A}}|^{\frac{1}{2}(n+p-r)}$, with $c^*$ a function of m, n, r, and p alone.

Using a well known determinental identity we may now write the complete density as

$$c''\left|[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\underline{\underline{H}}_{12}]\underline{\underline{\Omega}}_{22}[\underline{\underline{R}}_{1.2}-\underline{\underline{H}}_{11}^{-1}\underline{\underline{H}}_{12}]^{t}+\underline{\underline{H}}_{11}^{-1}\right|^{-\frac{1}{2}(n+p-r)}$$

where c'' is as defined in assertion 4 of section 2.